Submitted to *Adv. in Space Res,* Proceedings of the COSPAR Colloquium in Beijing China, 2002

Implementation and Verification of the Chen Prediction Technique for Forecasting Large Nonrecurrent Storms

C. Arge[1,2], S. Wahl[1,2], J. Chen[3], S. Slinker[3], and V. Pizzo[2]

[1]*Cooperative Institute for Research in Environmental Sciences, University of Colorado, Campus Box 216, Boulder, CO 80309, USA*
[2]*Space Environment Center, National Oceanic and Atmospheric Administration, R/SEC 325 S. Broadway, Boulder, CO 80305, USA*
[3]*Plasma Physics Division, Code 6790, Naval Research Laboratory, 4555 Overlook Ave. S.W., Washington, D.C., 20375, USA*

ABSTRACT

The Chen prediction technique is designed to identify and predict accurately the occurrence, duration, and strength of large geomagnetic storms using real-time solar wind data. It estimates the interplanetary magnetic field and the geoeffectiveness of the solar wind upstream of a monitor and can provide warning times that range from a few hours to (in principle) more than 10 hours. The model identifies physical features of solar wind structures that cause large storms: long durations of southward interplanetary magnetic field. It is currently undergoing testing, improvement, and validation at the Space Environment Center at the National Oceanic and Atmospheric Administration to transition it into a real-time space weather forecasting tool. In this paper, we report on the results of a 3-year historical verification study of the hourly-updated version of the model using ACE MAG Level 2 data. A real-time prediction web page has been developed and is on line at the Space Environment Center.

INTRODUCTION

It is well known that solar wind (SW) streams with large and sustained southward magnetic fields produce geomagnetic storms (e.g., Gonzalez and Tsurutani, 1987). Spacecraft located at L1 can in principle detect approaching geoeffective SW structures and provide forecasters with an advance warning time of about an hour, the SW transit time from L1 to the magnetosphere. While a number of existing models can successfully predict the geoeffectiveness of SW disturbances using L1 data (e.g., O'Brien and McPherron, 2000), currently only the Chen model (Chen *et al.,* 1996; 1997) can reliably predict the geoeffectiveness of approaching SW magnetic structures with warning times that significantly exceed the 1 hour SW transit time from L1 satellites. In this paper, we report on a comprehensive verification study conducted on the Chen technique for forecasting large nonrecurrent storms.

MODEL DESCRIPTION

The Chen model (Chen *et al.,* 1996; 1997) is a feature-based pattern recognition scheme that identifies and predicts the occurrence, duration, and strength of large geomagnetic storms using real-time SW data. The model estimates the interplanetary magnetic field (IMF) and the geoeffectiveness of the SW upstream of an L1 monitor. The essence of the technique is to infer the overall magnetic structure of a storm-producing disturbance (i.e., long durations of southward IMF) upstream of the monitor by examining the physical properties of the initial segment. The model works because magnetic field properties within streams that produce storms differ statistically from those that do not (Chen *et al.,* 1997). It is useful for forecasting purposes because geoeffective streams can frequently be identified after observing as little as 20% of a storm producing event, which (due to their large scale sizes) can often yield warning times significantly greater than the ~1 hour provided by L1 monitors.

The model makes two basic assumptions about storm producing SW streams: 1) they represent large, magnetically organized structures and 2) the duration and severity of a storm is a function of both the strength and duration of a stream's southward IMF. The model thus looks for extended periods of southward B_z. Each hour, the algorithm, as currently implemented at SEC, retrieves the previous hour's SW data from an L1 monitor such as the Advanced Composition Explorer (ACE), adds it to previously retrieved data (top 3 panels in Figure 1), and then searches for the most recent sign change in B_z. Beginning at the sign change and using B_y and B_z, it calculates the average rate of change in the IMF clock angle and predicts the time when B_z next changes sign by assuming that the time variation is a sine curve (panel c of Figure 1). This yields the predicted duration (τ) of the current IMF event (i.e., defined by the model as any northward or southward B_z excursion from zero back to zero) is the time interval from the last B_z sign change to the next as determined by the sine curve profile (panel d of Figure 1). The predicted maximum B_z excursion (B_{zm}) (panel e of Figure1) of the event is the maximum value of the predicted sine curve. The predicted fraction of the event observed (ξ) is the time interval (Δt_z) from the last B_z sign change to the most recent B_z datum retrieved divided by the predicted duration of the event (i.e., $\Delta t_z/\tau$). The model uses these predicted parameters, along with statistical knowledge of storm producing SW magnetic structures and Bayes theorem (which revises the probability of an event in light of new information) to calculate the probability (panel f of Figure1) that the event will generate a storm exceeding a specified threshold with the estimated duration τ and maximum B_z. With each new set of hourly SW data, the predicted qualities are revised and a new storm probability is generated. The model is designed to identify moderately large to large storms as they have the greatest potential for causing disruption and/or damage to power grids, satellites, and communications. IMF events are therefore defined to be geoeffective if they cause the *Dst* index to fall below –80 nT for more than 2 hours. The statistical properties of past IMF disturbances that caused storms are used to create a statistical database in the form of probability density functions (PDFs) that the model then uses to generate storm probabilities (see Chen et. al, 1997 for details). To avoid verifying the model with the same set of SW data used to generate the PDFs, the model was verified using ACE data but the PDFs were generated using WIND satellite data.

ANALYSIS

In 2000, a preliminary verification study was conducted on the Chen model using 2-years of real-time ACE MAG data available at the National Oceanic and Atmospheric Administration/Space Environment Center (NOAA/SEC). While the results obtained were promising, a new verification study using the significantly more reliable ACE MAG Level 2 data (available at http://www.srl.caltech.edu/ACE/) was deemed necessary due to the problematic nature of real-time ACE data - especially those available in 1998. The goal here is to establish a reliable benchmark of the model's performance using the most reliable data possible with the understanding that its performance will be diminished, on occasion, from this ideal level due to the provisional nature of real-time data.

Three years (1998-2000) of ACE MAG Level 2 data were used to verify the model. During this time interval, there were 43 instances where the *Dst* index fell below –80 nT for more than 2 hours, but in two cases, no ACE data were available. Final (1998-1999) and provisional (2000) *Dst* index data were obtained from the World Data Center for Geomagnetism in Kyoto, Japan (http://swdcdb.kugi.kyoto-u.ac.jp). Of the 41 events where *Dst* and ACE data were both available, 33 (~80%) were successfully predicted, i.e., the storm probability generated by the model exceeded 50% at some point during or shortly after an IMF event. The average lead-time (the time interval from the end of the hour when the probability first exceeded 50% to the point where *Dst* fell below –80 nT) provided by the model was 2.1 hours with a standard deviation of 2.5 hours (i.e., ~67% of the successfully predicted events had lead-times that fell between 4.6 and 0 hours). For one event, a warning time of 10 hours was obtained, and in 8 cases, a lead-time of 0 hours was assigned because the *Dst* index fell below –80 nT before the storm probability exceeded 50% (i.e., there was no advanced warning of the event). On such occasions, the predictions made by the model are still useful because they can provide forecasters with an idea of the severity of the storm (i.e., *Dst* will remain below –80 nT for than 2 hours) and the expected duration and maximum excursion of southward B_z. Currently, the algorithm does not predict the length of warning time.

For the 33 successfully predicted storms, the average unsigned fractional deviation between the predicted and observed B_{zm} after 10%, 20%, and 30% of the way into the IMF events is, respectively, 0.4, 0.3, and 0.3. The corresponding values for τ are 0.8, 0.6, and 0.4. At the point when the alarm was triggered (i.e., when the probability first exceed 50%), the average unsigned fractional deviations for B_{zm} and τ are 0.2 and 0.5, respectively. The model is thus better at predicting B_z maximum of an event than its duration. One reason for this is that the model is reset whenever B_z changes sign - no matter how briefly. Many examples were found during southward IMF events where B_z briefly fluctuated from south to north through zero and then back thus resetting the model. Another reason for poorly predicted event durations is due to the simple assumption that such IMF events have sinusoidal shapes.

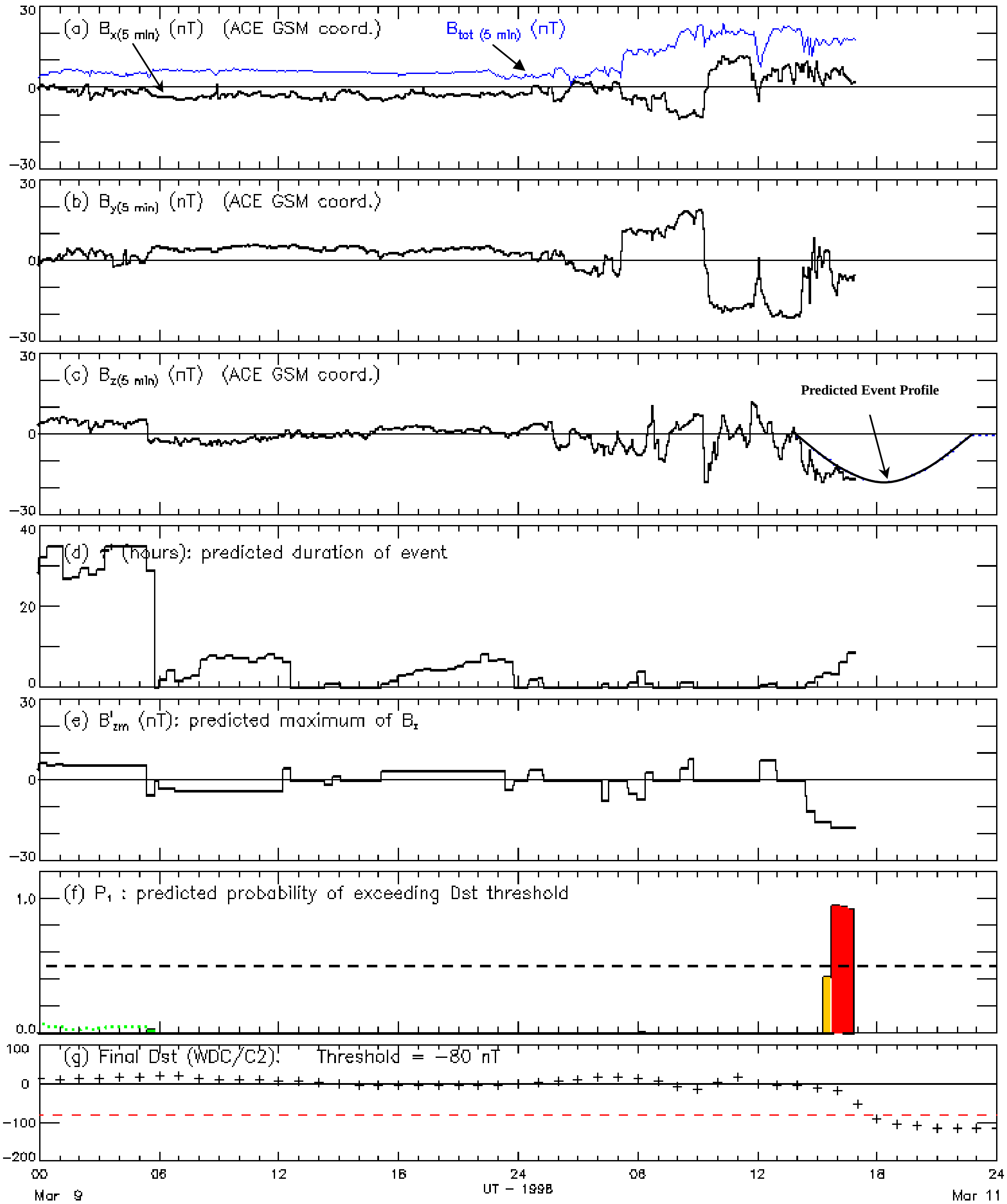

Fig. 1 – Example of a successful storm prediction made by the Chen model (see text for details).

The model missed 8 storms, i.e., *Dst* fell below –80 nT for more than 2 hours but the storm probabilities generated did not reach 50%. In 4 of these cases, the *Dst* index barely fulfilled the model's strict definition of a storm (i.e., these cases were marginal). Three misses can be explained because, in its present form, the model implicitly assumes that the *Dst* index lies at an undisturbed level before the arrival of a geoeffective event. In these 3 cases the *Dst* index was instead just above the –80 nT threshold when a minor B_z south excursion (one that probably would not have generated a storm on its own) succeeded in pushing the *Dst* index below –80 nT for 2 hours. In two-thirds of the misses, the IMF disturbances only had southward fields of moderate size but flow speeds exceeding 550 km s^{-1}, and the more quickly a moderately sized southward directed magnetic field is driven into the magnetosphere the more likely it is to generate a storm. Thus it is very likely that many of these storms were missed because the model does not presently take into account the flow speed of the disturbances. We plan to incorporate SW speed into the model in the near future. We note that the model successfully predicted all large storms and missed only those of moderate size, typically with (maximum) negative *Dst* above –100 nT.

There were 20 cases where the storm probability exceeded 50% but no storm subsequently occurred. These cases are divided into two categories: warnings and false alarms. Warnings are defined as cases where the model generates a storm probability that falls between greater than 50% and less than 70%, and false alarms are defined as cases where the probability is greater than or equal to 70%. In the former, the model's confidence that a storm will ensue is not very high, while in the latter, it is very confident that one will occur. If properly optimized, the model should be correct only half the time for a storm probability of 50%. A storm probability of 70% was selected as the cutoff between warning and false alarms because in all but one of the 33 successfully predicted events, the storm probability exceeded 70%. In 4 of the 9 cases where a warning occurred, the *Dst* index hovered very close to but did not fall below –80 nT. In one case, the (provisional) *Dst* index was exactly –80 nT for 2 consecutive hours but was classified as a warning because of strict adherence to definition. In other warning cases, the model predicted excessively long event durations ($\tau > 30$ hours) that eventually pushed the predicted probability over the 50% threshold even though the maximum B_z during these events was never very large. We note that the average event duration in the 33 successfully predicted storms was approximately 14 hours, with two cases having durations lasting upwards of 28 hours. Fitting the event profiles with a simple sine function is the likely source of these overly long duration predictions. In fact, a 35-hour duration cutoff was imposed to eliminate additional warnings and false alarms. The two major culprits producing many of the 11 false alarms were the excessively long duration predictions just mentioned and complicated event profiles that simply confused the model and resulted in erroneous τ and B_{zm} predictions.

Our results also appear to reveal the seasonal effect as described by Cliver et al. (2000) where the magnetosphere becomes less responsive to the SW near the solstices due to a reduction in coupling efficiency. Unpublished work by David Webb (private communication) shows evidence of this effect during solar cycle 23. Webb found that there are substantially fewer storm days (i.e., where *Dst* falls below –100 nT) during the 4 months near the solstices (i.e., December, January, June, and July) compared to the rest of the year. In apparent confirmation of these results, we found that ~50% of the false alarms and warnings made by the Chen model occur during these same four months, while one would expect only ~33% to occur due to chance. Only 10% of the correctly predicted events occur during the above four months.

An hourly updated real-time Chen model web page is currently available on line at NOAA/SEC at http://solar.sec.noaa.gov/~narge/cloud/cloud.cgi. It is presently being updated and will soon include an improved graphical display that includes real-time *Dst* predictions available from http://sprg.ssl.berkeley.edu/dst_index/ that can be directly compared with the Chen model predictions. It will also have model explanation and example pages along with suitable links to other web sites. The model has also been tested using historical WIND data. The test results and the SW data used for the test are archived at http://wwwppd.nrl.navy.mil/prediction/.

SUMMARY

Using 3-years (1998-2000) of ACE MAG Level 2 data, we have conducted a comprehensive verification study of the Chen prediction technique for forecasting large nonrecurrent storms. The model successfully predicted 33 out 41 (i.e., ~80%) storms for which there is concurrent ACE and *Dst* data with only 11 false alarms, 9 warnings, and 8 misses. For the 33 successfully predicted storms, it provided an average warning time of 2.1 hours with a maximum of ~10 hours and minimum of 0 hours. Our results also reveal the seasonal effect as described by Cliver *et al.* (2000).

ACKNOWLEDGEMENTS

We would like to thank ACE/MAG team members N. F. Ness, J. L'Heureux, and C. W. Smith at Bartol Research Institute and L. Burlaga, and M. Acuña at Goddard Space Flight Center for access to the ACE MAG Level 2 data. We would also like to thank the staff at the World Data Center for Geomagnetism/Data Analysis Center for Geomagnetism and Space Magnetism at Kyoto University, Japan for allowing us use of their hourly *Dst* index data. This work was funded by the Office of Naval Research under grant N00014-01-F-0026.

REFERENCES

Chen, J., P. J. Cargill, and P. J. Palmadesso, *Geophys. Res. Lett.*, **23**, 625, (1996).
Chen, J., P. J. Cargill, and P. J. Palmadesso, *J. Geophys. Res.*, **102**, 14701, (1997).
Cliver, E. W., Y. Kamide, and A. G., Ling, *J. Geophys. Res.*,**105**, 2413, (2000).
Gonzalez, W. D, and B. T. Tsurutani, *Planet. Space Sci.*, **35**, 1101, (1987).
O'Brien, T.P., and R.L. McPherron, *Journal of Atmospheric and Solar-Terrestrial Physics*, **62**, 1295, (2000).